http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

GEORGIA ALBUQUERQUE
georgia@cg.cs.tu-bs.de
Computer Graphics Lab, TU Braunschweig

DIRK J. LEHMANN
dirk@isg.cs.uni-magdeburg.de
Visual Computing Group, University of Magdeburg

TORBEN RODERMUND
rodermund@cg.cs.tu-bs.de
Computer Graphics Lab, TU Braunschweig

MARTIN EISEMANN
eisemann@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

THOMAS NOCKE
nocke@pik-potsdam.de
Potsdam Institute for Climate Impact Research

HOLGER THEISEL
theisel@isg.cs.uni-magdeburg.de
Visual Computing Group, University of Magdeburg

MARCUS MAGNOR
magnor@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Semi-Automatic Classification of Weather Maps

Technical Report 2012-3-17

March 20, 2012

Computer Graphics Lab, TU Braunschweig

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

Contents

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

Abstract

In this paper we analyze weather maps to distinguish between the three main circulation forms which are essential factors for weather composition and are fundamental for weather forecasters. We propose a set of features specifically tailored for the classification of these circulation forms in *General Weather Situations* and use these to train a support vector machine for classification. Additionally, we propose a semi-automatic algorithm to extract the necessary data directly from the weather maps itself. This enables us to also analyze historic map material for which the original data is not available anymore. In order to reconstruct the weather data, we extract and analyze the isolines from the weather maps based on color and line thickness as well as symbolic and numerical features using template matching techniques. We reconstruct the dense wind alignment field and air velocity field from these sparse data and extract expressive feature vectors to classify the presented main circulation forms. Our algorithm shows an overall classification success rate of 61% for the three main circulation forms zonal, meridional and mixed.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

1 Introduction

The circulation patterns of the atmosphere are an essential factor in weather composition and are fundamental for weather forecasters. Surface and upper air weather maps depicting positions for high and low pressure regions deliver a good approximation of the circulation. These maps are defined over a geographical region at a specified time, based on information from weather stations. For weather predictions, they help to classify the atmospheric state in *General Weather Situations* (GWS). F. Baur describes GWS as the average distribution of air pressure over a large region (at least the size of Europe) [WG09]. These *Weather Situations* can be subdivided into three main circulation forms, e.g. in Europe: zonal, meridional and mixed.

The decision of which *Weather Situation* prevails in the atmosphere is made by specialists that analyze the maps at different days (at least 3 days with the same patterns). A necessity for the classification is the knowledge of the air circulation in the atmosphere. The upper air weather maps depict positions for high and low pressure regions from which the circulation needs to be derived. Some earlier approaches [Lun63, MS85] classify weather situations by automatically defining a set of criteria or features. However, despite their automatism, the classification using the automatically extracted features might dissent from the subjective opinion of the specialist. In the work of James *et al.* [Jam07] a hybrid approach was therefore proposed which combined a manual feature selection with an objective classification scheme. Unfortunately, classification is only one part of the problem when analyzing weather maps. In the case of historic map material the necessary data needs to be reconstructed first. None of the aforementioned techniques is able to provide automatic classification of the circulation forms from the weather maps itself.

We propose to use a semi-automatic approach to reconstruct air circulation information from upper air weather maps to classify GWS. Our main idea is to first interactively extract the air pressure isolines from the weather maps. Template matching extracts the symbolic features, e.g. the high and low air pressure peaks, denoted by “H” and “T” respectively in Figure 2, as well as the numerical symbols depicting the absolute value of air pressure along the line. We reconstruct a dense wind alignment field and air velocity field and extract an optimized feature descriptor from these which in turn is used for supervised training of a support vector machine (SVM). This way the subjective knowledge of the specialist is incorporated opposed to using only objective features from the data.

Our approach can support meteorologists to understand the weather behaviour on large scales. Standard weather or distribution forecasts are some of the applications. Regarding accidents like the destruction of the Fukushima power plant, in Japan 2011, the GWS in such regions can steer in large scale how a radiation plume distributes in the atmosphere. Furthermore, for climate research, changes in long-term GWS are an important driver/indicator of regional climatic change, and support the understanding of local trends (e.g. of changes in temperature or

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

precipitation patterns). In such cases, the correct classification of the circulation form in GWS is of highest priority.

Additionally, our interactive reconstruction of air pressure from maps itself is beneficial to classify the weather situations in historic map material for which the original data is lost. The reconstructed information can then be used for data mining of weather data archives, e.g. to compare current conspicious constellations with similar situations and their outcome in the past.

In the rest of the paper we describe our algorithm in gradual steps. After reviewing previous work in Section 2, we explain our classification approach in Section 3, including our feature extraction and reconstruction of the wind alignment and air velocity field, Section 3.1, as well as our optimized feature descriptor and SVM classifier, Section 3.2. Section 4 conveys an analysis and discussion of our classification results. We conclude in Section 5 with a summary and give a prospective outlook on future work.

2 Related Work & Background

In this interdisciplinary work, we combine techniques from the field of visualization, image processing, and machine learning in order to process a popular issue within the field of meteorology. Therefore, our related work covers several research areas which will be discussed within this section. Furthermore, we will discuss the basic image data material which is the basis of our classification approach.

2.1 Classification of General Weather Situations

The research area of *synoptic scale meteorology* mainly deals with the classification of the GWSs. The European GWSs are subjectively classified by considering several characteristic features. Nevertheless, the used classification is not arbitrary but based on a history of experiences: The history of the European GWS classification has started at the beginning of the 19_{th} century with a simple classification of F. Baur with 21 GWS classes. This first classification distinguished mainly between the geographical position of the prediction, the position of the active front and the characteristic precipitation (cyclonic or anticyclonic). Later on, between 1946 and 1954, the synoptic scale meteorology rapidly developed, leading to the currently established classification of Paul Hess and Helmuth Brezowsky [HB52]. This last classification scheme was updated in subsequent works [HB77],[GW99] and [WG09]. The number of considered features has increased and encompasses now the distribution of the atmospheric pressure, the distribution of air masses or certain boundaries of those air masses, as well as the trajectories of their movements over the time. Additionally, considered features are the sunshine duration or the amount of precipitation. The most actual classification [GW99] distinguishes 29 classes of European GWS.

Similar classes are summarized in so-called General Weather Types (GWT)

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

distinguished by the main air flow direction (east-west, north-south and so on), and these GWTs are further distinguished by the kind of circulation: Altogether, there is the zonal circulation (4 classes of European GWS), meridional circulation (18 classes of European GWS), and mixed circulation forms (7 classes of European GWS). The classification of European GWS strongly depends on finding objective and reasonable features. Availability of such features and the ability of their detection is a limiting factor in an objective classification. In the next section we describe the image data material used in this work in more detail. These will serve as basis for the features which we will use for classification of the main circulation forms.

2.2 Properties of Weather Image Data

The underlying weather data for this work are image data. Figure 1 illustrates a typical example of a weather image data annotated with its containing features: The locations of the minima/maxima of a high/low pressure region are marked by “T” and “H”, respectively. A dashed/drawn through curve marks an isoline of constant air pressures (isobar), where the value of air pressure is given through a dedicated number on the isoline. The direction of the wind or air flow is approximately parallel to these isobars [WG09]. The air pressure difference between adjacent isolines is 4 hPa. In addition, for each image data the corresponding period of time is given and the underlying geographic region (Europe) is presented in the background.

Besides those direct features, also some indirect features can be derived. The underlying air flow field is uniquely orientated because the direction of the air flow is counterclockwise for a low pressure region and clockwise for a high pressure region, in the northern hemisphere. Moreover, the Coriolis force based on the rotation of the earth that refracts a straight air flow, is already intrinsicly coded by the behavior of the isolines.

From fluid dynamics it is known (cf. Bernoulli’s principle) that the difference within the local hydrostatic air pressure $|\nabla p|$ (estimated by its gradient) is proportional to the local velocity v of the underlying fluid (neglecting the friction and compression-based density fluctuation): $|\nabla p| \approx v$. Due to that, the slope of the air pressure between adjacent isobars is 4; it follows that the closer the isobars are to each other, the larger the difference of the air pressure is. This enables us to estimate $|\nabla p|$, and further, we can also estimate the local air flow velocity from this. In other words: the density distribution of the isobars approximately codes the local air flow velocity.

We will use those features of the image data to generate an appropriate classification of GWS. The question follows how a classification can be automatically generated. The area of machine learning addresses this issue, as will be discussed in the next sections.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

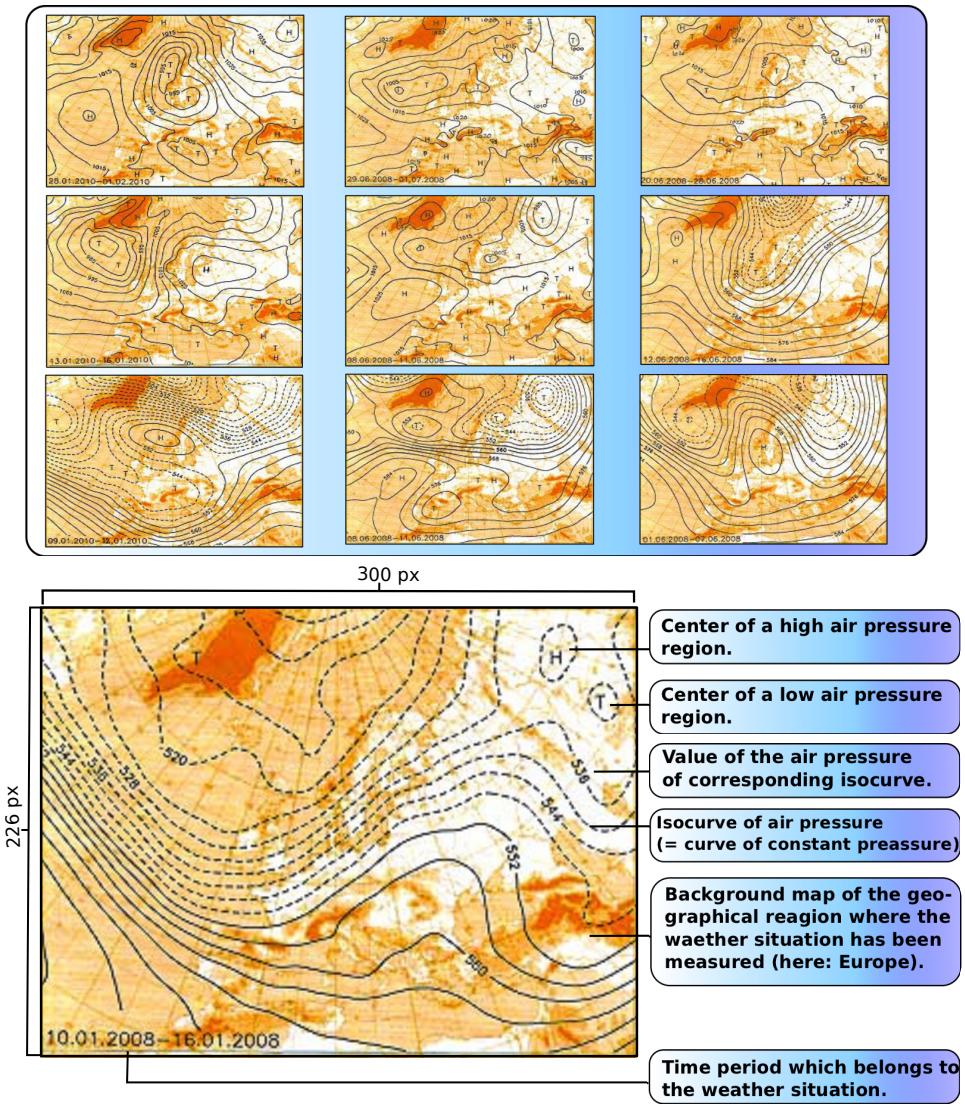


Figure 1: Available image material for analyzing the classification of weather data: (up) exemplary weather image data showing different *General Weather Situations*, (down) annotation of the general features within the image data.

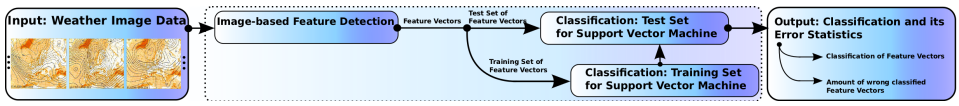


Figure 2: Workflow of the image data-based classification.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

3 Our Approach

Figure 2 illustrates our workflow for classification of the weather maps. First, each weather map has to be converted into a single fixed feature vector (Section 3.1). These features are used as input for the SVM, and to ensure a proper classification, the vector components should represent essential attributes to differ the classes (Section 3.2). In this step, the feature vectors are splitted into two disjoint sets: a training set that is used to create the classification model and a test set with unseen examples to test the efficiency of the trained model.

3.1 Interactive Feature Extraction

This section describes the feature extraction of image data. The weather map images used in this work were provided by specialists from the *German Weather Service* and the *Potsdam Institute for Climate Impact Research*. Some of these maps have a low resolution of 300×226 pixel and due to the bad image quality (see Figure 1) there are some handicaps for the feature extraction process: The pressure values, represented by three integers numbers cannot not be precisely extracted. The numbers are too small and in some cases illegible. Multiple digitalizations from analog images caused a heavy quality loss. In other cases, pressure symbols (Hs and Ts) cover each other. Thus, due to the resolution limitation of some weather maps, we focus on rather sharp and clear features like isobars [PS07], the alignment of the air flow and the date of acquisition. Our approach automatically derives the features from the image data.

An image function is given by

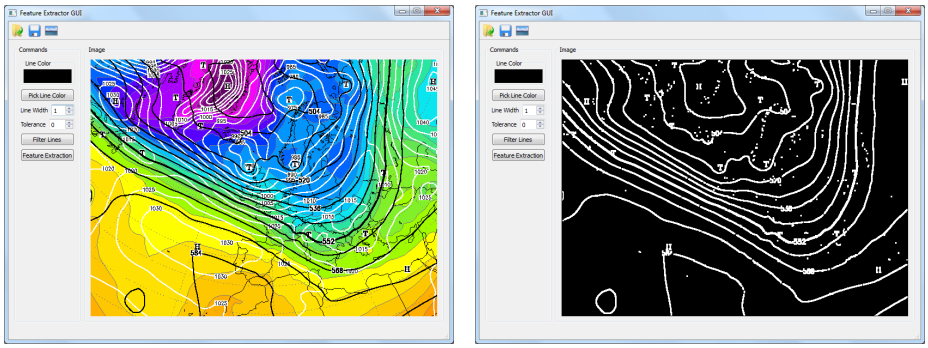
$$\mathbf{I}(x,y) \rightarrow (r(x,y), g(x,y), b(x,y)) : \mathbb{N}^2 \rightarrow \mathbb{R}^3$$

whereas $r,g,b \in [0,1] \in \mathbb{R}$ are the three color channels of each pixel $\mathbf{I}(x,y)$ concerning the 2D position $(x,y)^T \in [0, x_{max}-1] \times [0, y_{max}-1]$.

Interactive foreground extraction . The weather maps are composed of the foreground (lines, numbers and pressure symbols) and an underlying map of Europe. The main focus is the foreground which has to be separated for further processing. We provide a user interface to support the background subtraction in different kind of maps with different colors, scales and backgrounds. Figure 3 shows an example of foreground extraction. First, the user selects the color of the isolines (by picking it with the mouse) and then defines how large the lines are allowed to be, defining the *Line Width* in pixels. Finally, a *Tolerance* can be adjusted, to include pixels with similar colors to the selected one in the foreground. Once we have defined these parameters, they can be automatically used to process several maps with the same properties. Figure 3(b) shows the final foreground for this example.

In the next step, lines, numbers and symbols have to be extracted from the binary map. Therefore pixel components have to be classified into sets. A recursive

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012



(a) (b)

Figure 3: Preparing the feature extraction by separating foreground: (a) Weather map from DWD, (b) Computed foreground.

function can easily detect connected pixels to form a list of components. Due to the bad resolution, features of different kinds might be contained in a single group. This can happen to pressure symbols which are too close to line strips. Afterwards the approach calculates the connected components for each group of adjacent pixels to prepare feature extraction. Figure 4 (a-b) illustrates this.

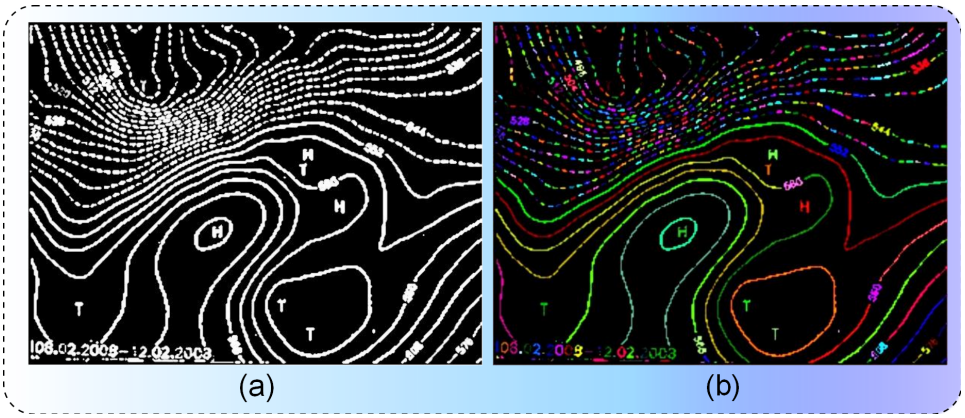


Figure 4: Preparing feature extraction by identifying components: (a) Example of a binary image, (b) connected group of adjacent pixels.

Classification of connected components Components can be classified into template symbols (*symbol features*, e.g., “H” and“T”), into numerical symbols (*number features*, e.g., “680”) or into *line features* like dashed or solid lines.

Template symbols are detected by pixel masks which are grid-based template matchings [Bru09]. The mask is moved across each component and evaluated at all positions. Similar to convolution each position returns a value which represents the

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

matching quality. If all source pixels fit into the pixel mask, there will be a matching pattern at the current position. The □ entry is an arbitary pixel which does not affect the matching quality. Detection is getting more stable by using wildcards.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & \square & 1 & \square & 0 \\ 0 & \square & 1 & \square & 0 \\ 0 & \square & 1 & \square & 0 \\ 0 & \square & 1 & \square & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & \square & \square & \square & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & \square & \square & \square & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The classification of *number features* is done by bounding boxes. All numbers on a given weather map appear to have 3 digits and the same size. As the font is almost monospaced[1], a bounding box can handle this kind of feature pretty well. The numbers are embedded into the line strip of isobars. Therefore the orientation is changing from place to place. To classify an arbitary component into *number features*, a box of fixed size[2] is rotated around the barycenter. With each iteration, the angle $\alpha \in [0, 2\pi]$ is increased. If all pixels of the current component are inside the box at a certain rotation, it might be a number feature. To improve the results of this test, the pixels of the component should cover a certain percentage of the corresponding box. About $70-80\%$ seems to be a good approximation of needed pixel density in *number features*.

Some of the *number features* are splitted into several components, e.g., when 3 digits are not connected (680 is separated into 68 and 0). Here we need to combine close components and have to repeat the test. Our approch fuses up to 2 components to detect separated *number features*.

Based on this classification we can separate *symbol features* and *numbers features*. In the following we will explain the extraction of isolines, wind alignment fields and wind velocity fields.

Isoline Reconstruction Let $B(x,y)_{LF}$ be the binary image, which only contains those connected components that are classified as *line features*. All the connected components in $B(x,y)_{LF}$ correspond to isobars. Some of the isobars are unfortunately presented as dashed lines (Figure 5 (a)). Therefore, the approach has to reconstruct isolines by connecting dashes to solid lines.

Common approaches to detect dashed lines by applying morphological operations [BBE07] have a lack in line density. As soon as the fragments of different lines are too close and in parallel order, these operations start melting fragments. Therefore, a more sophisticated approach is needed.

[1] Font which is having a fixed width for each character

[2] The size of box is based on our experience with the weather image data.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

At first the fragments of the dashed lines are distinguished from the solid lines. In general dashes have a rather small number of pixels, therefore we can classify them as "dots", which have a number of pixels $l \leq l_{dot}$ with l_{dot} is the 10% quantile of the histogram $H(l(S(x,y)_{LF}))$ given by:

$$l_{dot} := \int_0^{l_{dot}} H(l(S(x,y)_{LF}))\, dl = 0.1.$$

Figure 5 (b) illustrates this. For practical reasons an appropriate approximation of l_{dot} is given by: $l_{dot} = l_{max} \cdot 0.05$.

For each line fragment the topological skeleton [GW01] is computed which represents the further alignment (principal component) of the fragment. To determine if two fragments can be connected, the facing endpoints of their skeletons are compared. Each endpoint has a direction vector based on the remaining skeleton pointing outside. The angle between the facing vectors is a metric for the connection quality of a pair of endpoints. Sequenced fragments of the same line are having small angles. Fragments which do not belong to the same line have a greater value. With this angle we can find partners and their connection quality in the surrounding neighbourhood. They are only combined if both fragments will appoint each other as best candidates by not exceeding the metric. This process is repeated by decreasing the threshold for the best candidate. At first, the fragments have to point to each other without any tolerance. By increasing the tolerance, more and more fragments are combined forming proper line strips. It is important to increase the tolerance of admissible candidates slowly to avoid parallel connection of different lines.

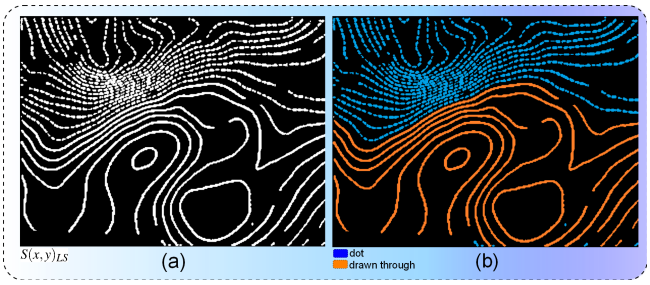


Figure 5: Isobar reconstruction: (a) example of binary image $S(x,y)_{LF}$ with (b) classified dots and solid lines.

Wind Alignment Field Another feature which is important for the classification of general weather situations is the principle alignment of the air flow. The air alignment field $\mathbf{f}(x,y) = (u(x,y), v(x,y))^T : \mathbb{R}^2 \to \mathbb{R}^2$ is a certain vector and flow field, respectively. For the reconstructed isobars $S(x,y)_{LF}$, our algorithm approximates the gradient ∇S_{LF} by a convolution based on the Sobel operator [KS11] G_x

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

and G_y:

$$\nabla S_{LF} = \begin{pmatrix} S(x,y)_{LF}/\partial x \\ S(x,y)_{LF}/\partial y \end{pmatrix} = \begin{pmatrix} G_x * S(x,y)_{LF} \\ G_y * S(x,y)_{LF} \end{pmatrix};$$

$$\text{with } G_x = G_y^T \text{ and } G_y = \frac{1}{8}\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.$$

For each pixel $(x_p,y_p)^T$, which is on an isobars in $S(x,y)_{LF}$, the normalized air alignment field $\mathbf{f}(x_p,y_p) = (u(x_p,y_p),v(x_p,y_p))^T$ is given by:

$$\mathbf{f} = \begin{pmatrix} \cos(\tan^{-1}(\beta)) \\ \sin(\tan^{-1}(\beta)) \end{pmatrix}; \beta = \begin{cases} \frac{S(x_p,y_p)_{LF}/\partial y}{S(x_p,y_p)_{LF}/\partial x}; S(x_p,y_p)_{LF}/\partial x \neq 0 \\ \frac{\pi}{2}; S(x_p,y_p)_{LF}/\partial x, S(x_p,y_p)_{LF}/\partial y = 0 \\ 0; \; else \end{cases}.$$

For all pixels (or sub-pixels) which are not on the isobars in $S(x,y)_{LF}$, the air flow field $\mathbf{f}(x,y)$ is calculated by a Shepard-Interpolation [She68] based on the air alignment field $\mathbf{f}(x_p,y_p)$ for pixels on the isobars:

$$\mathbf{f} = \begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \frac{((x-x_{p_i})^2+(y-y_{p_i})^2)^{-1}}{\sum_{j=1}^{n}((x-x_{p_j})^2+(y-y_{p_j})^2)^{-1}} \cdot u(x_{p_i},y_{p_i}) \\ \sum_{i=1}^{n} \frac{((x-x_{p_i})^2+(y-y_{p_i})^2)^{-1}}{\sum_{j=1}^{n}((x-x_{p_j})^2+(y-y_{p_j})^2)^{-1}} \cdot v(x_{p_i},y_{p_i}) \end{pmatrix},$$

with n being the number of pixels on the isobars (cf. Figure 6).

Air Velocity Field: As discussed in Section 2.2, hydrostatic pressure and air velocity are directly caused and related(under reasonable assumption) by the same effect: the mass conservation of fluid mechanics. Therefore there are actually two possibilities to use the air pressure information and the gradient of the air pressure, respectively. The approach either reconstructs the pressure scalar field, similar to [HSS03], or the air velocity scalar field. Using both possibilities as feature is not reasonable, because they have different aspects of the same information and are consequently redundant.

We decided to use the approach that reconstructs the air velocity field because this can be done very quickly and efficiently. The air velocity field $v(x,y) : \mathbb{R}^2 \to \mathbb{R}$ is a scalar field of the local velocity air flow.

As already mentioned in Section 2.2, the density distribution of the isobars represents the local air flow velocity. Therefore, the approach estimates the corresponding density distribution $v(x,y)$ via an iteratively applied convolution with a Gaussian G_σ convolution kernel [HA91] to isobars of $S(x,y)_{LF}$:

$$v(x,y) = G_\sigma * S(x,y)_{LF}, \text{ with } G_\sigma(m,n) = \frac{1}{2\pi\sigma^2} e^{-\frac{m^2+n^2}{2\sigma^2}}$$

The idea is to interpret the isobars themselves as density functions with a density of 1 and to blur the density into the neighbourhood so that regions with many isobars have a larger density in average (= velocity) than other regions. Figure 6 summarizes the feature extraction results.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

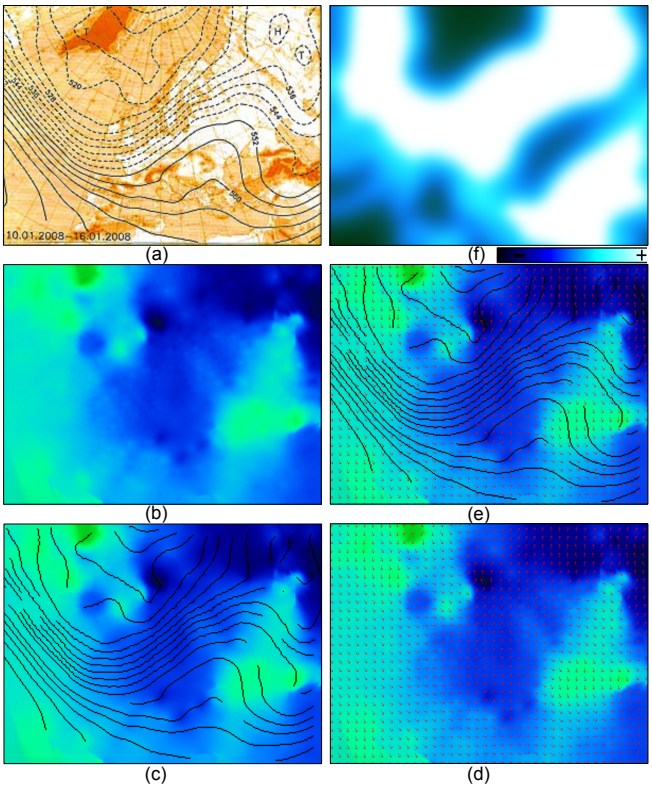


Figure 6: Results of the automatic feature extraction from the weather image data: (a) original image data, (b) reconstructed air flow field (color coded $u(x,y)/v(x,y)$ is the red/green channel), (c) air flow field with reconstructed isobars, (d) air flow field (vector field illustration), (e) air flow field (vector field illustration) with isobars, (f) air velocity field (color coded).

3.2 Feature-based Classification of GWS

Once we have reconstructed an approximation of the weather maps, one of the most challenging issues in the weather map classification is the decision of which set of features should be used to represent the maps. According to the catalog of the *General Weather Situations of Europe* proposed by Hess and Brezowsky [HB77], there are main subjective features that are used to manually classify the circulation forms (zonal, meridional or mixed). The zonal circulation form, e.g., is characterized by a subtropical high-pressure region over the North Atlantic and a system of low pressure in the subpolar area, creating a rather straight west-to-east flow between them, where single low-pressure regions move with their front-line from west to east, i.e. from the eastern part of the North Atlantic to the European continent. On the other side, the meridional circulation form can be characterized by an existence of stationary and blocking high-pressure regions between 50 and 65 degrees of northern latitude. Also, depressions with north-south axis direction

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

are considered to be of this circulation form. In a mixed circulation the zonal and meridional wind components have almost the same proportion. Based on this information, our vector feature is composed by: the month (from 1 to 12) on which the map was created, the wind direction and magnitude in main regions of the map. These main regions of the map are defined by a grid. We concentrate in the central region of the image and discard the four corners of the image because they do not have impact on the results. The size of the feature vector depends on the grid resolution. By subdividing each region the classification error can be influenced. We evaluated our classification framework the same set of weather maps with different subdivisions and different vector configurations (see Figure 7). The horizontal axis describes how many subdivisions were done on the central region of the map and the vertical axis depicts the classification error. Figure 8 shows three examples of grid resolution that can be used. Figure 8 (a) shows a grid with only 1, (b) with 2 and (c) with 5 subdivisions.

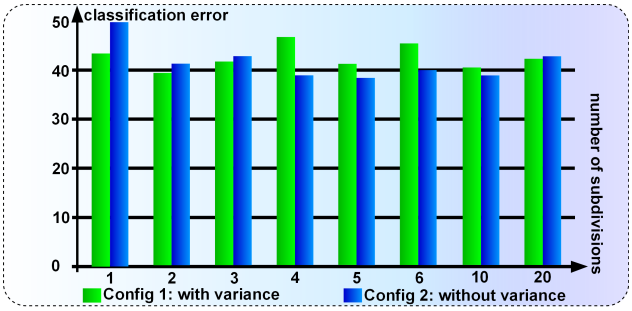
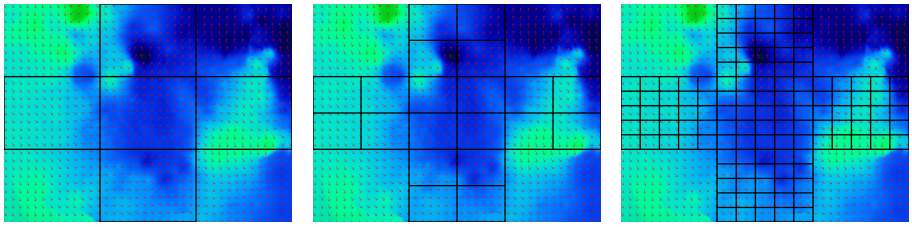


Figure 7: Different grid tilings and vector configurations and the associated error rate. Config1 considers the variance. Config2 does not consider the variance.

The best results were delivered by not using the variance (Config2) and the resolution schema represented in Figure 8(c) , i.e. 5 subdivisions. With a feature vector containing 40 wind elements that are computed per grid cell as the mean of the 90% quantile of all pixels in the cell for each wind component (x and y). These wind components are defined by the direction (see Figure 6(d)) increased by the wind magnitude (see Figure 6(f)).

The last step of our framework is the final classification which assigns one of the defined labels to the feature vectors. As aforementioned, we categorize the weather maps in three different classes, i.e., the circulation forms. We adopted a classification approach that follows a geometric approach based on decision boundaries. We use a multiclass extension of the Support Vector Machine [CS02] (SVM) in order to categorize the selected features. The original SVM proposed by Vapnik [Vap98] supported only binary decisions, it was later extended to a multilabel variant by Crammer and Singer [CS02]. This multiclass approach can be used in cases where more than two classes are defined. We use this extension to train our classifier for the three circulation forms. After a short training period, a prediction model is created and can be used to predict the class of new unseen weather maps.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012



(a) (b) (c)

Figure 8: Different resolutions of the grid to compute the wind features. The scheme represented in (c) delivered the best results.

The used training and test sets are described in detail in Section 4.

4 Results

For our classification approach, 603 weather image maps were available. Altogether, these maps describe the European GWS over the time period from January 2002 to December 2010. To validate our classification framework, we partitioned the maps into two distinct sets. We used 423 (70 %) images as training set and 180 (30 %) as test set. The classes in the training and test set are not well distributed and only a low number of examples of the *Zonal* class was available in both sets. For the training set e.g., the *Zonal*, *Mixed* and *Meridional* classes had 68, 178 and 176 samples, respectively. Similarly, for the test set, each of the classes have 34, 68 and 78 samples, respectively.

Using the features described in Section 3.2 and in Figure 8(c), our classification framework achieved an overall classification success of 61.67% for the test set and the three classes. This result is far from a random selection of 33.33% for three classes. The individual results per class can be seen in Table 1. Note that the *Zonal* class presented a very low percentage of correctness in the classification of 17.65%, while the other two classes *Mixed* and *Meridional* presented much better results: 61.76% and 80.77% of the samples were correctly classified. The low classification rate for the *Zonal* class may occur because the Zonal class was underrepresented in the training set, with only 68 samples, i.e., 15% of the total samples. Therefore, not all possible configurations for the *Zonal* class are represented in training set. However, most misclassified samples from *Zonal* and *Meridional* classes were categorized as *Mixed*. This result is quite intuitive because the *Mixed* circulation form has elements of the other both (*Zonal* and *Meridional*). In this case the misclassification is not a serious failure because a map classified as *Mixed* can be further evaluated by weather specialists in a interactive process.

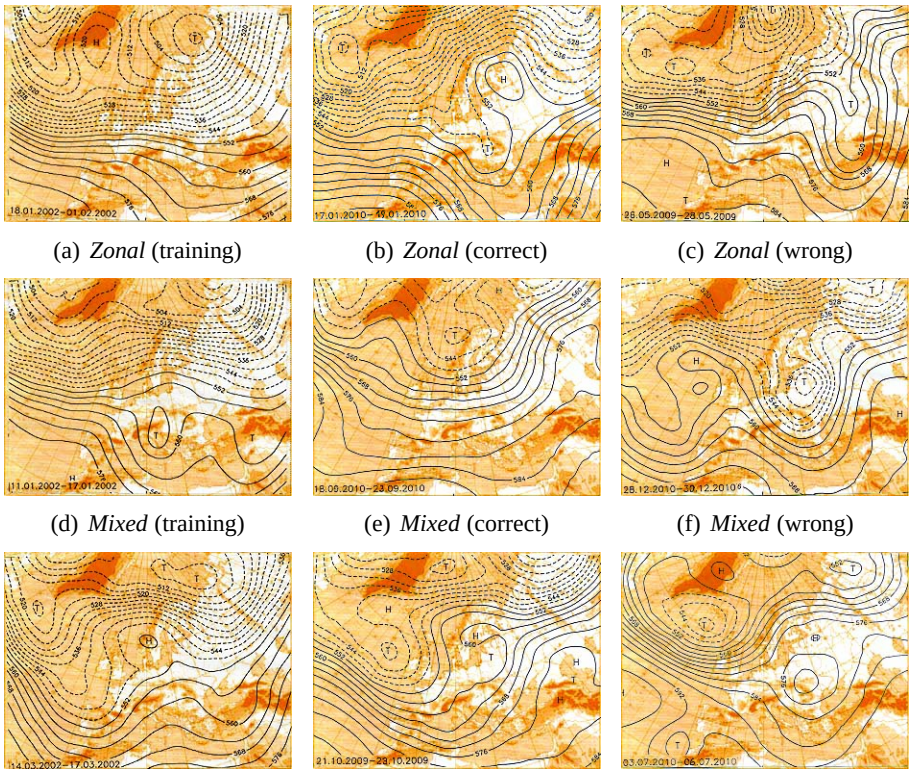
Figure 9 shows examples of weather maps that belong to the three different categories. In the first column are examples of the training set for the three classes, in the second are examples of the test set that were successfully classified by our framework, while the last column presents examples of the test set that were mis-

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

classified. Note that the visual classification of such maps in GWS is not trivial and can be very difficult for non-specialists.

Table 1: Classification Results

Classes	Zonal	Mixed	Meridional
Zonal	17.65%	52.94%	29.41%
Mixed	4.41%	61.76%	33.82%
Meridional	12.83%	17.95%	80.77%



(a) *Zonal* (training) (b) *Zonal* (correct) (c) *Zonal* (wrong)

(d) *Mixed* (training) (e) *Mixed* (correct) (f) *Mixed* (wrong)

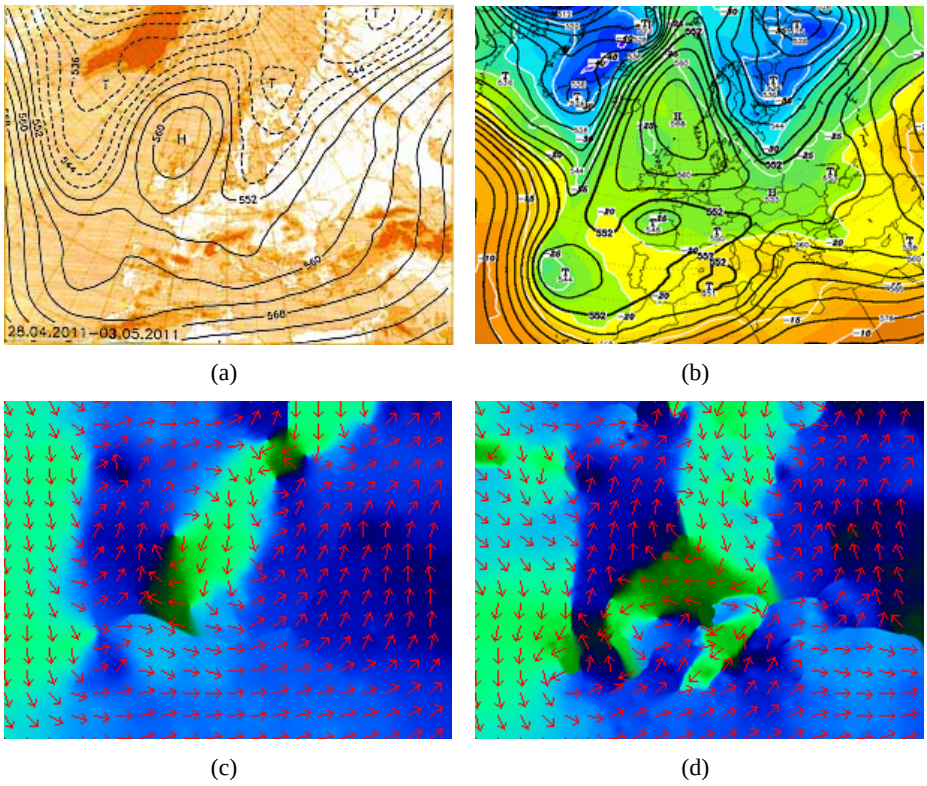
(g) *Merid.* (training) (h) *Merid.* (correct) (i) *Merid.* (wrong)

Figure 9: Examples of weather maps from our data sets. The first column (a, d and g) shows examples of the training set for the three distinct classes, the second column shows examples of the test set that were successfully classified and the last column presents examples of the test set that were misclassified. (c) is known as *Zonal* but was classified as *Meridional*, (f) is known as *Mixed* but was classified as *Meridional* and (i) is known as *Meridional* but was classified as *Zonal*.

As aforementioned, the proposed framework supports several map formats. Figure 10 shows an example with a different map type. This second type is similar to the first one, with a different background and a better resolution of 640×472.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

The two maps belong to the *Meridional* circulation form and were correctly classified by our system. Figures 10 (a) and (b) show the original maps and Figures 10 (c) and (d) the respective air flow fields.



(a) (b)

(c) (d)

Figure 10: Classification example with different maps. Both maps belong to the *Meridional* circulation form and were correctly classified by our system. (a) and (b) show the original maps and Figures 10 (c) and (d) the respective air flow fields.

5 Conclusion

In this work, we proposed a classification framework to support the categorization of weather maps in General Weather Situations (GWS). To the best of our knowledge, this is the first work that establishes a workflow for such task only based on weather map images. Our framework brings together established techniques used in the area of visualization and image processing and machine learning. We describe a set of features from weather map images that are relevant for this classification according to the Hess and Brezowsky main circulations forms, and we present a plausible method for the extraction of these features. Moreover, we trained a classification model using support vector machines to validate our features. In our first experiments we were able to classify weather maps in the three

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

pre-defined circulation forms with an overall accuracy rate of 61% percent. Our classification results can be used as a pre-categorization to guide the final classification by weather specialists.

As future work we intent to test our framework with other feature sets and larger data sets to improve its classification rate. A larger number of training samples, mainly for the Zonal class can significantly improve the results. Moreover, based on this extended training set, we intent to classify not only the circulation forms, but also the different 29 possible *European General Weather Situations*.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

References

[BBE07] Gerald J.F. Banon, Junior Barrera, and Ulisses M. Braga-Neto (Eds.). Mathematical morphology and its applications to signal and image processing. *Proceedings of the 8'th international symposium on mathematical morphology*, 2007.

[Bru09] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley, 2009.

[CS02] Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.*, 2:265–292, March 2002.

[GW99] F.-W. Gerstengarbe and P.C. Werner. *Catalog of the General Weather Situations of Europe 1881 - 1998 referring to Paul Hess und Helmuth Brezowsky / Katalog der Großwetterlagen Europas 1881 - 1998 nach*. German Meteorological Service / Deutscher Wetterdienst, 1999.

[GW01] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2001.

[HA91] R.A. Haddad and A.N. Akansu. A class of fast gaussian binomial filters for speech and image processing. *Signal Processing, IEEE Transactions on*, 39(3):723 – 727, 1991.

[HB52] P. Hess and H. Brezowsky. *Catalog of the General Weather Situations of Europe / Katalog der Großwetterlagen Europas*. German Meteorological Service / Deutscher Wetterdienst, 1952.

[HB77] P. Hess and H. Brezowsky. *Catalog of the General Weather Situations of Europe 1981-1976 / Katalog der Großwetterlagen Europas 1881-1976*. German Meteorological Service / Deutscher Wetterdienst, 1977.

[HSS03] K. Hormann, S. Spinello, and P. Schröder. C1-continuous terrain reconstruction from sparse contours. In *VMV*, pages 289–297, 2003.

[Jam07] P. M. James. An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe. *Theoretical and Applied Climatology*, 88(1):17–42, January 2007.

[KS11] P. R. Kumar and K.L. Sailaja. Watermarking algorithm using sobel edge detection. *International Journal of Advanced Networking and Applications*, 2:861–867, 2011.

[Lun63] I.A. Lund. Map pattern classification by statistical methods. *Journal of Applied Meteorology*, 2:56–65, 1963.

http://www.digibib.tu-bs.de/?docid=00042812 21/03/2012

[MS85] R.H. Maryon and A.M. Storey. A multivariate statistical model for forecasting anomalies of 1=2-monthly mean surface pressure. *Journal of Climatology*, 5:561–578, 1985.

[PS07] J. Pouderoux and S. Spinello. Global contour lines reconstruction in topographic maps. In *Proceedings of ICDAR 2007: 9th International Conference on Document Analysis and Recognition*, pages 779–783. IEEE Computer Society, sep 2007.

[She68] Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, ACM '68, pages 517–524, New York, NY, USA, 1968. ACM.

[Vap98] V. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, New York, 1998.

[WG09] P.C. Werner and F.-W. Gerstengarbe. *Catalog of the General Weather Situations of Europe 1881-2009 (7. Edition) / Katalog der Großwetterlagen Europas 1881-2009 (7. Auflage)*. German Meteorological Service / Deutscher Wetterdienst, 2009.